REGULATION OF EMBRYONIC TRANSCRIPTION IN PLANTS

FIELD OF THE INVENTION

The invention is in the field of nucleic acid sequences capable of regulating transcription, particularly sequences that may promote transcription during embryogenesis in plants.

BACKGROUND OF THE INVENTION

Most of the information about seed-specific gene expression comes from studies of genes encoding seed storage proteins like napin, a major protein in the seeds of *Brassica napus*, or conglycinin of soybean. Upstream DNA sequences directing strong embryo-specific expression of these storage proteins have been used successfully in transgenic plants to manipulate seed lipid composition and accumulation (Voelker et al., 1996). However, expression of storage protein genes begins fairly late in embryogenesis. Thus, promoters of seed storage protein genes may not be ideal for all seed-specific applications. For example, storage oil accumulation commences significantly before the highest level of expression of either napin (Stalberg et al., 1996) or conglycinin (Chen et al., 1988) is achieved. It is, therefore of interest to identify other promoters which may modulate expression of genes in developing plant embryos.

A variety of transcriptional regulatory regions that may be active during plant embryogenesis are known, as disclosed for example in: U.S. Patent No. 5,792,922 issued 11 August 1998 to Moloney; U.S. Patent No. 5,623,067 issued 22 April 1997 to Vandekerckhove et al.; International Patent Publication WO9845461 published 15 October 1998. There remains a need for alternative transcriptional regulatory regions.

FATTY ACID ELONGATION1 (*FAE1*) genes encode condensing enzymes involved in plant very long chain fatty acid biosynthesis. The FAE1 condensing enzyme is thought to be localized in the endoplasmic reticulum where it catalyzes the sequential elongation of C18 fatty acyl chains to C22 in length (Kunst et al., 1992). *FAE1* genes have been cloned and described recently by James et al. (1995), International Patent Publication WO 96 13582.

SUMMARY OF THE INVENTION

In one aspect, the invention provides transcriptional regulatory regions derived from *FAE1* genes. The transcriptional regulatory regions of the invention may be useful in

promoting early seed-specific transcription of heterologous sequences to which they are operably linked. The transcriptional regulatory regions of the invention may be used in a wide variety of plants, including *Brassica sp.*, *Arabidopsis* and other plant species. DNA constructs comprising the transcriptional regulatory sequences of the invention may be active during fatty acid or lipid biosynthesis in the plant embryo. Certain embodiments of the constructs of the invention may be used in transgenic plants to promote expression of heterologous sequences in developing seeds. In various embodiments, the constructs of the invention may be used to mediate gene expression that affects seed lipid metabolism, or seed protein composition or seed carbohydrate composition, or seed development. In alternative embodiments, the transcriptional regulatory regions of the invention may also be useful for the production of modified seeds containing novel recombinant proteins which have pharmaceutical, industrial or nutritional value.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a 934 bp DNA sequence comprising the *Arabidopsis thaliana FAE1* transcription regulatory sequence.

Figure 2 shows a 1588 bp DNA sequence comprising the *Brassica napus FAE1* transcription regulatory sequence.

Figure 3 shows a 1069 bp DNA sequence comprising the *Lunaria annua FAE1* transcription regulatory sequence.

Figure 4 shows an alignment of the *Arabidopsis thaliana* (*A.t.*), *Lunaria annua* (*L.a.*) and *Brassica napus* (*B.n.*) transcription regulatory sequences. Asterisks below the sequences indicate identical nucleotides in each of the three sequences. A number of putative cis-acting sequence motifs are identified in the *A. thaliana* sequence: an EM1 ABA box at -44bp to -36bp having the sequence ACATCTCAT, for which the published consensus sequence is ACGTGTCAT (Rowley, D.L. and Herman, E.M. (1997), Biochimica et Biophysica Acta 1345:1-4); an A-300 box at -51bp to -46bp having the sequence TGCAAT, for which the published consensus sequence is TG(T/A/C)AAA(G/T) (Morton et al. (1994) in Seed Development and Germination (Kigel, J. and Galili, G., eds.) pp. 103-138, Marcel Dekker, New York); G-box 1 at -105 to -100 bp having the sequence CACATG, for which is the consensus sequence is CACCTG, and G-box 2 at -164 to -159 bp having the sequence CAACTT, for which the consensus sequence is CAACTG (Kawogoe, Y. and Murai, N. (1992) Plant J. 2:927-936; CE1 element at -226 to -218 bp having the sequence

TTCCATCGA, for which the consensus sequence is TGCCACCGG, and a CE3 element at -381bp to -369 bp having the sequence ACACATTCCCTC, for which the consensus sequence is ACGCGTGTCCTC (Shen et al., (1996) Plant Cell 8:1107-1119). Not highlighted is a putative RY repeat motif at -53bp to -47bp having the sequence CATGCAA, for which the consensus sequence is CATGCAT (Dickinson et al. (1988) Nucleic Acid Res. 16:371; Lelievre et al. (1992) Plant Physiol. 98:387-391). Also shown, as Con. 4, is a consensus sequence, wherein R=G or A, Y=T/U or C, M=A or C, K=G or T/U, S=G or C, W=A or T/U, B=G or C or T/U, D=A or G or T/U, H=A or C or T/U, V=A or G or C and N=A or G or C or T/U.

Figure 5 shows an alignment of the *Arabidopsis thaliana* (*A.t.*) and *Lunaria annua* (*L.a.*) transcription regulatory sequences. Asterisks below the sequences indicate identical nucleotides in each of the two sequences. The base at position -400 in the *A.t.* sequence is highlighted. The alignment of sequences in both Figure 4 and Figure 5 was accomplished using the CLUSTALW program (version 1.74) for multiple sequence alignments, using a gap open penalty of 15, a gap extension penalty of 6.66 and an IUB DNA weight matrix. Also shown, as Con. 5, is a consensus sequence, wherein R=G or A, Y=T/U or C, M=A or C, K=G or T/U, S=G or C, W=A or T/U, B=G or C or T/U, D=A or G or T/U, H=A or C or T/U, V=A or G or C and N=A or G or C or T/U.

Figure 6 includes two bar graphs illustrating hydroxy fatty acid content of A) *FAE1-FAH12* and B) *napin-FAH12* transgenic seeds, expressed as percentage of total seed fatty acids.

Figure 7 shows an alignment of the *Brassica napus* (*B.n.*) and *Lunaria annua* (*L.a.*) *FEA1* transcription regulatory sequences. Asterisks below the sequences indicate identical nucleotides in each of the two sequences.

Figure 8 shows an alignment of the *Brassica napus* (*B.n.*) and *Arabidopsis thaliana* (*A.t.*) *FEA1* transcription regulatory sequences. Asterisks below the sequences indicate identical nucleotides in each of the two sequences.

DETAILED DESCRIPTION OF THE INVENTION

The recombinant nucleic acid molecules of the invention may comprise a heterologous promoter sequence operably linked to a nucleic acid sequence, wherein the promoter sequence comprises a transcriptional regulatory region capable of mediating seed-specific expression in *Arabidopsis*. The transcriptional regulatory region may be obtainable from a

plant *FAE1* gene. Alterntively, The transcriptional regulatory region may hybridize under stringent conditions to a 5' region of the plant *FAE1* gene. In further alternative embodiments, The transcriptional regulatory region may be at least 70% identical when optimally aligned to the 5' region of the plant *FAE1* gene.

In alternative embodiments, the invention provides isolated nucleic acids comprising the transcriptional regulatory regions of the invention. By isolated, it is meant that the isolated substance has been substantially separated or purified away from other biological components with which it would otherwise be associated, for example *in vivo*. The term 'isolated' therefore includes substances purified by standard purification methods, as well as substances prepared by recombinant expression in a host, as well as chemically synthesized substances.

In the context of the present invention, "transcriptional regulatory region" means a nucleotide sequence capable of mediating or modulating transcription of a nucleotide sequence of interest, when the transcriptional regulatory region is operably linked to the sequence of interest. Conversely, a transcriptional regulatory region and a sequence of interest are "operably linked" when the sequences are functionally connected so as to permit transcription of the sequence of interest to be mediated or modulated by the transcriptional regulatory region. In some embodiments, to be operably linked, a transcriptional regulatory region may be located on the same strand as the sequence of interest. The transcriptional regulatory region may in some embodiments be located 5' of the sequence of interest. In such embodiments, the transcriptional regulatory region may be directly 5' of the sequence of interest or there may be intervening sequences between these regions. The operable linkage of the transcriptional regulatory region and the sequence of interest may require appropriate molecules (such as transcriptional activator proteins) to be bound to the transcriptional regulatory region, the invention therefore encompasses embodiments in which such molecules are provided, either *in vitro* or *in vivo*.

The term "recombinant" means that something has been recombined, so that when made in reference to a nucleic acid molecule the term refers to a molecule that is comprised of nucleic acid sequences that are joined together by means of molecular biological techniques. The term "recombinant" when made in reference to a protein or a polypeptide refers to a protein molecule which is expressed using a recombinant nucleic acid molecule. The term "heterologous" when made in reference to a nucleic acid sequence refers to a nucleotide sequence which is ligated to, or is manipulated to become ligated to, a nucleic acid sequence

to which it is not ligated in nature, or to which it is ligated at a different location in nature. The term "heterologous" therefore indicates that the nucleic acid molecule has been manipulated using genetic engineering, i.e. by human intervention.

Sequences may be derived or obtainable from plant *FAE1* genes by deduction and synthesis based upon the wild-type *FAE1* gene sequences. Derived sequences may be identified in different organisms, for example by isolation using as probes the nucleic acid sequences of the invention. Alternative transcriptional regulatory regions may be derived through mutagenesis or substitution of wild-type sequences, such as the sequence disclosed herein. Derived nucleic acids of the invention may be obtained by chemical synthesis, isolation, or cloning from genomic DNAs using techniques known in the art, such as the Polymerase Chain Reaction (PCR). Consensus sequences, such as those illustrated in Figures 4 and 5 are alternative embodiments of the nucleic acids of the invention, derived from the disclose wild-type *FAE1* gene sequences. Nucleic acids of the present invention may be used to design alternative primers (probes) suitable for use as PCR primers to amplify particular regions of an *FAE1* gene. Such PCR primers may for example comprise a sequence of 15-20 consecutive nucleotides of the sequences of the invention. To enhance amplification specificity, primers of 20-30 nucleotides in length may also be used. Methods and conditions for PCR amplification are described in Innis et al. (1990); Sambrook et al. (1989); and Ausubel et al. (1995). As used herein, the term "probe" when made in reference to an oligonucleotide refers to an oligonucleotide which is capable of hybridizing to another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are, for example, useful in the detection, identification, amplification and isolation of particular gene sequences. Oligonucleotide probes may be labelled with a "reporter molecule," so that the probe is detectable using a detection system, such as enzymatic, fluorescent, radioactive or luminescent detection systems.

Derived nucleic acids of the invention may also be identified by hybridization, such as Southern or Northern analysis. Southern analysis is a method by which the presence of DNA sequences in a target nucleic acid mixture are identified by hybridization to a labeled probe, comprising an oligonucleotide or DNA fragment of a nucleic acid of the invention. Probes for Southern analysis may for example be at least 15 nucleotides in length. Southern analysis typically involves electrophoretic separation of DNA digests on agarose gels, denaturation of the DNA after electrophoretic separation, and transfer of the DNA to nitrocellulose, nylon, or another suitable membrane support for analysis with a radiolabeled, biotinylated, or enzyme-

labeled probe as described in Sambrook *et al.* (1989). Similarly, Northern analysis may be used to identify RNAs that hybridize to a known probe such as an oligonucleotide, DNA fragment, cDNA or fragment thereof, or RNA fragment of a nucleic acid of the invention or a known *FAE1* sequence. The probe may be labeled with a radioisotope such as ^{32}P, by biotinylation or with an enzyme. The RNA to be analyzed may be electrophoretically separated on an agarose or polyacrylamide gel, transferred to nitrocellulose, nylon, or other suitable membrane, and hybridized with the probe, using standard techniques well known in the art such as described in Sambrook *et al.* (1989).

In alternative embodiments, a transcriptional regulatory region of the invention may be at least 70% identical when optimally aligned to the 5' region of a plant *FAE1* gene, such as the *Arabidopsis FAE1* gene. In alternative embodiments, the degree of identity may be between 50% and 100%, such as 60%, 80%, 90%, 95% or 99%. When a position in the compared sequence is occupied by the same nucleotide or amino acid, following optimal alignment of the sequences, the molecules are considered to have identity at that position. The degree of identity between sequences is a function of the number of matching positions shared by the sequences. In terms of percentage, identity is the sum of identical positions, divided by the total length over which the sequences are aligned, multiplied by 100.

Various aspects of the present invention encompass nucleic acid or amino acid sequences that are homologous to other sequences. As the term is used herein, an amino acid or nucleic acid sequence is "homologous" to another sequence if the two sequences are substantially identical and the functional activity of the sequences is conserved (for example, both sequences function as or encode a FAE1 enzyme; as used herein, the term 'homologous' does not infer evolutionary relatedness). Nucleic acid sequences may also be homologous if they encode substantially identical amino acid sequences, even if the nucleic acid sequences are not themselves substantially identical, a circumstance that may for example arise as a result of the degeneracy of the genetic code.

Two amino acid or nucleic acid sequences are considered substantially identical if, when optimally aligned (with gaps permitted), they share at least about 50% sequence similarity or identity, or if the sequences share defined functional motifs. In alternative embodiments, sequence similarity in optimally aligned substantially identical sequences may be at least 60%, 70%, 80%, 90% or 95%. As used herein, a given percentage of homology between sequences denotes the degree of sequence identity in optimally aligned sequences.

Optimal alignment of sequences for comparisons of similarity may be automated using a variety of algorithms, such as the local homology algorithm of Smith and Waterman (1981) *Adv. Appl. Math* 2: 482, the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, the search for similarity method of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, and the computerized implementations of these algorithms (such as GAP, BESTFIT, FASTA and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, Madison, WI, U.S.A.). Sequence similarity may also be determined using the BLAST algorithm, described in Altschul *et al.* (1990), *J. Mol. Biol.* 215:403-10 (using the published default settings). Software and instructions for performing BLAST analysis may be available through the National Center for Biotechnology Information in the United States (including the programs BLASTP, BLASTN, BLASTX, TBLASTN and TBLASTX that may be available through the internet at http://www.ncbi.nlm.nih.gov/). The BLAST algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database (reference) sequence. T is referred to as the neighborhood word score threshold. Initial neighborhood word hits act as seeds for initiating searches to find longer HSPs. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction is halted when the following parameters are met: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLAST program may use as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (Henikoff and Henikoff (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919), a gap existence cost of 11, a per residue gap cost of 1, a lambda ratio of 0.85, alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands. One measure of the statistical similarity between two sequences using the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. In alternative embodiments of the invention, nucleotide or amino acid sequences are considered substantially identical if the smallest sum probability in a comparison of the test sequences is less than about 1, preferably less than about 0.1, more preferably less than

about 0.01, and most preferably less than about 0.001. In the PSI-BLAST implementation of the BLAST algorithm, an expect value for inclusion in PSI-BLAST iteration may be 0.001 (Altschul et al. (1997), Nucleic Acids Res. 25:3389-3402). Searching parameters may be varied to obtain potentially homologous sequences from database searches.

An alternative indication that two nucleic acid sequences are substantially identical is that the two sequences hybridize to each other under moderately stringent, or preferably stringent, conditions. Hybridization to filter-bound sequences under moderately stringent conditions may, for example, be performed in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65EC, and washing in 0.2 x SSC/0.1% SDS at 42EC (see Ausubel, *et al.* (eds), 1989, *Current Protocols in Molecular Biology*, Vol. 1, Green Publishing Associates, Inc., and John Wiley & Sons, Inc., New York, at p. 2.10.3). Alternatively, hybridization to filter-bound sequences under stringent conditions may, for example, be performed in 0.5 M $NaHPO_4$, 7% SDS, 1 mM EDTA at 65EC, and washing in 0.1 x SSC/0.1% SDS at 68EC (see Ausubel, *et al.* (eds), 1989, *supra*). Hybridization conditions may be modified in accordance with known methods depending on the sequence of interest (see Tijssen, 1993, *Laboratory Techniques in Biochemistry and Molecular Biology -- Hybridization with Nucleic Acid Probes*, Part I, Chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York). Generally, stringent conditions are selected to be about 5EC lower than the thermal melting point for the specific sequence at a defined ionic strength and pH.

A *FAE1* promoter is any naturally occurring transcriptional regulatory region that mediates or modulates the expression of a plant FAE1 condensing enzyme. Plant FAE1 condensing enzymes are proteins that are homologous to known FAE1 condensing enzymes, such as those cloned and described in International Patent Publication WO 96/13582.

Heterologous DNA sequences may for example be introduced into a host cell by transformation. Such heterologous molecules may include sequences derived from the host cell species, which have been isolated and reintroduced into cells of the host species. Heterologous nucleic acid sequences may become integrated into a host cell genome, either as a result of the original transformation of the host cells, or as the result of subsequent recombination events. Transformation techniques that may be employed include plant cell membrane disruption by electroporation, microinjection and polyethylene glycol based transformation (such as are disclosed in Paszkowski *et al. EMBO J.* 3:2717 (1984); Fromm *et al., Proc. Natl. Acad. Sci. USA* 82:5824 (1985); Rogers *et al., Methods Enzymol.* 118:627

(1986); and in U.S. Patent Nos. 4,684,611; 4,801,540; 4,743,548 and 5,231,019), biolistic transformation such as DNA particle bombardment (for example as disclosed in Klein, *et al.*, *Nature* 327: 70 (1987); Gordon-Kamm, *et al.* "The Plant Cell" 2:603 (1990); and in U.S. Patent Nos. 4,945,050; 5,015,580; 5,149,655 and 5,466,587); *Agrobacterium*-mediated transformation methods (such as those disclosed in Horsch *et al. Science* 233: 496 (1984); Fraley *et al., Proc. Nat'l Acad. Sci. USA* 80:4803 (1983); and U.S. Patent Nos. 4,940,838 and 5,464,763).

Standard methods are available for the preparation of constructs for use in identifying and characterizing transcriptional regulatory regions useful in various embodiments of the invention. General molecular techniques may for example be performed by procedures generally described by Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Stuhl K. (1995) Current Protocols in Molecular Biology, Vols 1, 2 and 3. Alternative equivalent methods or variations thereof may be used in accordance with the general knowledge of those skilled in this art and the functional requirements of the present invention.

In some aspects of the invention, transformed plant cells may be cultured to regenerate whole plants having a transformed genotype and displaying a desired phenotype, as for example modified by the expression of a heterologous protein mediated by a transcriptional regulatory region of the invention. A variety of plant culture techniques may be used to regenerate whole plants, such as are described in Gamborg and Phillips, "Plant Cell, Tissue and Organ Culture, Fundamental Methods", Springer Berlin, 1995); Evans *et al.* "Protoplasts Isolation and Culture", Handbook of Plant Cell Culture, Macmillian Publishing Company, New York, 1983; or Binding, "Regeneration of Plants, Plant Protoplasts", CRC Press, Boca Raton, 1985; or in Klee *et al., Ann. Rev. of Plant Phys.* 38:467 (1987). A cell, tissue, organ, or organism into which has been introduced a foreign nucleic acid, is considered "transformed", "transfected", or "transgenic". A transgenic or transformed cell or organism also includes progeny of the cell or organism and progeny produced from a breeding program employing a transgenic plant as a parent in a cross and exhibiting an altered phenotype resulting from the presence of a recombinant nucleic acid construct. A transgenic plant is therefore a plant that has been transformed with a heterologous nucleic acid, or the progeny of such a plant that includes the transgene. The invention provides vectors, such as vectors for transforming plants or plant cells. The term "vector" in reference to nucleic acid molecule generally refers to a molecule that may be used to transfer a nucleic acid segment(s) from one cell to another. One of skill will recognize that after the nucleic acid is stably incorporated in transgenic plants and

confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques may be used, depending upon the species to be crossed.

In various embodiments, the invention comprises plants transformed with the nucleic acids of the invention. In some embodiments, such plants will exhibit altered fatty acid content in one or more tissues. These aspects of the invention relate to all higher plants, including monocots and dicots, such as species from the genera *Fragaria, Lotus, Medicago Onobrychis, Trifolium, Trigonelia, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Hyoscyamus, Lycopersicon, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Cichorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Heterocallis, Nemesia, Pelargonium, Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browallia, Glycine, Lolium, Zea, Triticum, Sorghum, and Datura*. Such plants may include maize, wheat, rice, barley, soybean, beans, rapeseed, canola, alfalfa, flax, sunflower, cotton, clover, lettuce, tomato cucurbits, potato carrot, radish, pea lentils, cabbage, broccoli, brussel sprouts, peppers, apple, pear, peach, apricot, carnations and roses. More specifically, in alternative embodiments, plants for which the invention may be used in modifying fatty acid content include oil crops of the *Cruciferae* family: canola, rapeseed (*Brassica* spp.), crambe (*Crambe* spp.), honesty (*Lunaria* spp.) lesquerella (*Lesquerela* spp.), and others; the *Compositae* family: sunflower (*Helianthus* spp.), safflower (*Carthamus* spp.), niger (*Guizotia* spp.) and others; the *Palmae* family: palm (*Elaeis* spp.), coconut (*Cocos* spp.) and others; the *Leguminosae* family: peanut (*Arachis* spp.), soybean (*Glycine* spp.) and others; and plants of other families such as maize (*Zea* spp.), cotton (*Gossypium* sp.), jojoba (*Simonasia* sp.), flax (*Linum* sp.), sesame (*Sesamum* spp.), castor bean (*Ricinus* spp.), olive (*Olea* spp.), poppy (*Papaver* spp.), spurge (*Euphorbia*, spp.), meadowfoam (*Limnanthes* spp.), mustard (*Sinapis* spp.) and cuphea (*Cuphea* spp.).

Nucleic acids of the invention may also be used as a plant breeding tool, as molecular markers to aid in plant breeding programs. Such techniques would include using the gene itself as a molecular probe or using the DNA sequence to design PCR primers to use PCR based screening techniques in plant breeding programs.

Deletion or insertion constructs may be useful for domain mapping to determine the functional domains or motifs of a transcriptional regulatory region derived from a *FAE1* gene. An aspect of the invention is the construction and testing of such constructs, as described below for the 5' deletion construct of the *A. thaliana FAE1* 5' region. One aspect of the

invention comprises transcriptional regulatory regions that are derived from functionally important regions of a *FAE1* promoter. As outlined above, the functionally important regions of a *FAE1* promoter may be determined through routine assays. Alternatively, randomly selected portions of a a *FAE1* promoter may be selected for use in routine assays to determine whether the selected region is capable of functioning as a transcriptional regulatory region in the context of the present invention. In various embodiments, regions of the *Arabidopsis thaliana, Brassica napus* or *Lunaria annua* promoters may be used. For example, the following motifs in the *A.t. FAE1* promoter may be used alone or in combination in novel transcriptional regulatory regions (see Figure 4): the CE-like elements (CE1 and CE3), the RY repeat motif, the G-boxes (G-box1 and G-box2), the A-300 box, the EM1 ABA box, or the CTATTTTG element. Constructs of the invention comprising such motifs, deletions or insertions may be assayed for activity as transcriptional regulatory regions of the invention by testing for strong seed-specific activity providing expression of a sequence of interest (such as a reporter sequence) before the torpedo stage and persisting throughout embryo development, in accordance with standard testing methods that may be adapted from the methods disclosed herein.

Alternative embodiments of the transcriptional regulatory regions of the invention may be identified using information available through NCBI databases at http://www.ncbi.nih.gov.

In various embodiments, transcriptional regulatory regions derived from plant *FAE1* genes are shown to be capable of directing expression of desired genes at an early stage of development in a seed-specific manner in disparate plant species. In particular embodiments, the transcriptional regulatory regions of the invention may be used in a wide variety of dicotyledonous plants for modification of the seed phenotype. For example, new seed phenotypes may include:

(1) altered seed fatty acid composition or seed oil composition and accumulation

(2) altered seed protein or carbohydrate composition or accumulation

(3) enhanced production of desirable endogenous seed products

(4) suppression of production of undesirable gene products using antisense, co-suppression or ribozyme technologies

(5) production of novel recombinant proteins for pharmaceutical, industrial or nutritional purposes

Isolation of a seed-specific promoter from *A. thaliana*

Using the sequence information of the *A. thaliana* genome sequencing project, synthetic oligonucleotide primers were designed to amplify the *FAE1* 5' untranslated region, to isolate it by PCR. As shown in Figure 1, the upstream primer 5'-CTAGTAGATTGGTTGGTTGGTTTCC-3' (AtproFW) in combination with the downstream primer 5'-TGCTCTGTTTGTGTCGGAAAATAATGG-3' (AtproRV) were used, and resulted in the synthesis of a fragment of the correct size (934 bp). The amplified product was subcloned in the *Hinc*II site of the plasmid pT7T3-18U (Pharmacia) to produce plasmid pT7T3-18U/proFAE900, followed by complete sequence determination of both strands to verify the fragment identity. A BLAST search of the *A. thaliana* Database identified a single BAC clone T4L20 (GenBank ATF10M6) 125,179 bp long, which contains the complete *FAE1* gene.

Functional analysis of the *FAE1* 5' upstream region

5' upstream fragments of the *FAE1* gene were shown to confer seed-specific and temporally regulated gene expression in plants. A translational fusion was made between the *FAE1* 5' region and the coding region of the reporter gene β-glucuronidase (GUS). The chimeric gene (pFAE900-GUS or pFAE400-GUS) was transferred into *Arabidopsis* and tobacco and GUS activity was monitored in various tissue of transgenic plants.

Construction of the vectors pFAE900-GUS and pFAE400-GUS, and transformation of *Arabidopsis* and tobacco, was as follows. The insert was cleaved out of pT7T3-18U vector with *Hin*dIII and *Xba*I and directionally subcloned into the corresponding sites of the binary Ti plasmid pBI101 (Clontech), which contains a promoterless GUS gene (Jefferson et al. 1987), to obtain the vector pFAE900-GUS. Another construct, pFAE400-GUS, containing only 393 bp of the 5' *FAE1* region directly upstream of the ATG initiation codon fused to the GUS coding sequence was also generated. For that, the pT7T3-18U/proFAE900 vector was digested with *BglII* and *PstI*, the sticky ends were filled in using T4 DNA polymerase, followed by re-ligation to obtain pT7T3-18U/proFAE400. The 393 bp 5' *FAE1* upstream fragment was then excised with *Hin*dIII and *Xba*I and cloned into the binary vector pBI101 to obtain the plasmid pFAE400-GUS. The pFAE400-GUS and pFAE900-GUS fusion constructs in pBI101 were introduced into *Agrobacterium tumefaciens* strain GV3101 (Konez and Schell, 1986) by heat-shock and selected for resistance to kanamycin (50 μg ml). *A. thaliana* (L.) Heynh. ecotype Columbia was transformed with the pFAE400-GUS and pFAE900-GUS constructs using floral dip method (Clough and Bent, 1998). Screening for transformed seed

was done on 50μg/mL kanamycin as described previously (Katavic et al., 1994). Approximately 100 transgenic lines were generated for each construct.

For transformation of tobacco, *A. tumefaciens* harbouring the pFAE900-GUS construct was co-cultivated with leaf pieces of *Nicotiana tabacum* SR1 and transformants were selected with kanamycin (100μg/mL) on solid medium (Lee and Douglas, 1996).

Histochemical localization of GUS activity and analysis of transgenic plants was as follows. Tissue sections were placed in 100 mM $NaPO_4$ (pH7) and 1 mM spermidine for 15 min, then incubated at 37° C in 0.5 $K_3[Fe(CN)_6]$, 0.01 % Triton X-100, 1mM EDTA, 10 mM β-mercaptoethanol, 5-bromo-4-chloro-3-indolyl-β-D-glucuronide in 100 mM $NaPO_4$ (pH7), until a blue color appeared (after approximately 1 hr). Following incubation with the substrate, chlorophyll was removed from the sections using a graded ethanol series.

Using this assay, five independent transgenic *Arabidopsis* lines were examined for the embryo-specific expression of the GUS gene. In addition, leaf, stem and siliques were histochemically stained for β-glucuronidase activity. The results indicate that the reporter gene fused to the transcriptional regulatory region of the invention is not expressed in vegetative tissues, whereas it is highly expressed in developing seeds (embryos). Both the 934 bp and the 393 bp transcriptional regulatory regions derived from the *A.t. FAE1* gene caused the appearance of GUS activity by the torpedo stage embryo (6 days after flowering). GUS activity in all five lines persisted throughout subsequent embryo development.

Leaves, stems, pods and seeds of three regenerated tobacco lines transformed with the pFAE900-GUS construct were also assayed for β-glucuronidase activity. The results obtained indicate that the 934 bp *FAE1* promoter fragment contains sufficient information to direct seed-specific expression of a reporter gene in transgenic tobacco. Thus the transcriptional regulatory regions of the invention may be used for seed-specific expression of foreign genes in transgenic plants.

The *in vivo* activity of a *FAE1* promoter of the invention was compared to the activity of the napin promoter by expressing the castor bean hydroxylase gene *FAH12* (Broun and Somerville, 1997) behind either the *FAE1*-promoter (a transcriptional regulatory region of approximately 1 kb) or the napin promoter in an Arabidopsis *tad2 tae1* double mutant. This mutant accumulates as a proportion of fatty acids about 85% of the 18:1 acyl group, which is the substrate for the hydroxylase. The levels of hydroxylated fatty acids accumulating in a large number of independent transgenic lines were used to estimate the relative strength of

each promoter. As shown in Figure 6, the two populations of transgenic plants accumulated levels of hydroxylated fatty acids, ranging from 0.2% to about 11-12% of total fatty acids, with the levels being on average slightly higher in *FAE1-FAH12* lines. Similarly, the best *FAE1-FAH12* plant accumulated just over 12% of hydroxylated fatty acids (w/w of total FAs), whereas the best *napin-FAH12* plant produced 10.8% of hydroxylated fatty acids (w/w of total FAs). These results indicate that the FAE1 promoter is highly active in transgenic *Arabidopsis* and that its *in vivo* activity may be superior to napin in *Arabidopsis* seeds.

Sequence elements or motifs that confer both tissue specificity and developmental regulation of transcription reside within 393 bp of the AUG translation initiation codon in the *A.t. FAE1* gene. The seed-specific expression conferred by the transcriptional regulatory regions of the invention is independent of the native terminator of the *FAE1* gene 3' end. For example, in the exemplified constructs disclosed herein, a terminator derived from the *Agrobacterium* nopaline synthase gene was used.

Lunaria annua and *Brassica napus FAE1* 5'regulatory regions

Two sequences originating from *B. napus* and *L. annua* were isolated and characterized to demonstrate that regulatory regions conferring seed-specific transcription early in embryo development can also be found upstream of other plant *FAE1* genes. Sequences were cloned using the technique of polymerase chain reaction (PCR) walking on uncloned plant genomic DNA (Devic et al., 1997). Approximately 5 µg of genomic DNA from 1 g of fresh tissue was used for the construction of 5 different libraries by digesting DNA with a series of enzymes that produce blunt end fragments to which special adaptors are ligated. The adaptor molecules consist of a long upper strand, which contains successive sequences common to the adaptor primers, AP1 and AP2, annealed at its 3' end to a shorter strand lacking the AP1 sequence. However, this short strand posseses an amine group at its 3' end to prevent filling in by the DNA polymerses during the first PCR amplification step and generation of the AP1 binding site. This suppression PCR effect prevents exponential amplification of molecules containing the adaptor at each end, and the adaptor primer binding sites are only produced when a strand complementary to the upper strand of the adaptor is synthesized by extension from a gene specific primer. The first PCR reaction is performed using an adaptor primer AP1 and a gene specific primer. An aliquot of the first PCR product is used a template in a second PCR amplification using the nested gene specific primer and AP2.

In order to isolate the regulatory regions upstream of the *B. napus FAE1* coding sequence, genomic DNA was prepared from developing leaves and digested with 5 blunt-end

cutting restriction enzymes (*DraI, EcoRV, HpaI, PvuII and ScaI*) to generate a series of DNA libraries. After ligation of adapter molecules, individual libraries were used as templates in a two step PCR. In the first PCR amplification using the AP1 primer 5'-GGATCCTAATACGACTCACTATAGGGC-3' and the *FAE1* gene specific primer 5'-AAAGAGTGGAGCGATGGTTATGAGG-3' (Bnwalk1), multiple DNA fragments were amplified from all five library templates. After a second round of PCR, using the AP2 primer 5'-CTATAGGGCTCGAGCGGC-3' and the nested *FAE1* specific primer 5'-CGGAAAGAAGCAAAGGTTGAAAAGG-3' (Bnwalk2), the longest single fragment of 1.6 kb was obtained from the *HpaI* library template. This fragment was inserted into the pCR2.1 plasmid (Invitrogen) and sequenced. The sequence is shown in Figure 2.

For the PCR walking experiment to isolate the *L. annua* 5' regulatory region, in adition to the standard AP1 and AP2 primers, the following *FAE1* specific primers were used: 5'-GATCGTTTGTGGTAAGACGAGAGC-3' (Lawalk1) and 5'-GTCAGTGGGAAGAAACAGAGGTTG-3' (Lawalk2). In the first PCR reaction, the *DraI, EcoRV, PvuII, ScaI* and *SspI* library templates were used. In a second PCR amplification the longest single fragment 1.1 kb in length was synthesized using the *EcoRV* library template. This fragment was inserted into the *HincII* site of the pT7T3-18U vector (Promega), sequenced on both strands and analyzed (Figure 3).

Using the sequence data obtained for the 5' regulatory regions generated by PCR walking, specific primers were generated for the amplification of the *L. annua* and *B. napus FAE1* promoter fragments. For the PCR-amplification of *B. napus* promoter fragment the upstream primer was 5'-CTGACTTCACCAAAGAAACAACTCG-3' (BnproFW) in combination with the downstream primer 5'-CGGAATTCCGTTTTTTTTTTTAGGCG-3' (BnproRV). The synthesized fragment was ligated into the *SmaI* site of pGEM-7Zf (Promega), then excised with *XbaI BamHI* and cloned into the equivalent sites of the pBI101 binary vector (Clontech). *L. annua* 5' regulatory region was amplified using the 5'-CAGCTTAACCGGTAAAATTGGCC-3' (LaproFW) upstream primer together with the 5'-TGTTCAGTTTTGTGTCGGAGAGG-3' (LaproRV) downstream primer and inserted in the *HincII* site of pT7T3-18U (Promega) plasmid. In order to clone the *L. annua* promoter fragment into the pBI101 binary vector, an *XbaI* site was added by subcloning the *PstI KpnI* fragment released from the pT7T3-18U vector into pBluescript II KS+ (Stratagene). The fragment was then excised and cloned in the *XbaI* site of the pBI101 vector.

The resulting vectors pBnFAE1-GUS and pLaFAE1-GUS in pBI101 were then introduced into *A. tumefaciens* strain GV3101 by heat-shock, and used to transform *Arabidopsis* as described above. Transformants were selected on agar-solidified medium containing kanamycin (50 µg/ml). More than 100 transformants were generated for each construct. The activity of the *L. annua* and *B. napus FAE1* promoters was determined by GUS expression assays on the developing seeds and also on non-reproductive plant tissues as controls. Consistent seed-specific GUS expression was obtained for both promoter constructs in independent transgenic lines. In contrast, there was no detectable GUS activity in leaf, stem and silique samples.

References

The following documents are hereby incorporated by reference:

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Stuhl K. (1995) Current Protocols in Molecular Biology, Vols 1, 2 and 3.

Benfey, P.N. and Chua, N.-H. (1989) Regulated genes in transgenic plants. Science 244, 174-181.

Bevan, M. Shufflebottom, D., Edwards, K., Jefferson, R. and Schuch, W. (1989) Tissue- and cell-specific activity of a phenylalanine ammonia-lyase promoter in transgenic plants. EMBO. J. 8, 1899-1906

Broun, P. and Somerville, C. (1997) Accumulation of ricinoleic, lesquerolic, and densipolic acids in seeds of transgenic Arabidopsis plants that express a fatty acyl hydroxylase cDNA from castor bean. Plant Physiol. 113, 933-942

Chen , Z.L., Pan, N.S., and Beachy, R.N. (1988) A DNA sequence element that confers seed-specific enhancement to a constitutive promoter. *EMBO J.* 6: 3559-3564.

Clough,S.J. and Bent,A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabdiopsis thaliana*. *Plant J.* 16: 735-743.

Devic, M., Albert, S., Delseny, M., and Roscoe, T.J. (1997) Efficient PCR walking on plant genomic DNA. *Plant Physiol Biochem.* 35: 331-339.

James, D.W.,Jr., Lim, E., Keller, J., Plooy, I., Ralston, E., and Dooner, H.K. (1995) Directed tagging of the Arabidopsis *FATTY ACID ELONGATION (FAE1)* gene with the maize transposon *Activator*. *Plant Cell* 7: 309-319.

Jefferson, R.A., Kavanaugh, T. and Bevan, M.W. (1987) GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker system in higher plants. *EMBO J.* 6: 3901-3907.

Katavic, V., Haughn, G.W., Reed, D., Martin, M., and Kunst, L. (1994) *In planta* transformation of *Arabidopsis thaliana*. *Mol.Gen.Genet.* 245: 363-370.

Koncz, C. and Schell, J. (1986) The promoter of T_L-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol. Gen. Genet.* 204: 383-396.

Kunst, L., Taylor, D.C., and Underhill, E.W. (1992) Fatty acid elongation in developing seeds of *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 30: 425-434.

Lee D., and Douglas C.J. (1996) Manipulation of plant gene expression using antisense RNA. In: *Plant Biochemistry/Molecular Biology Laboratory/Manual*. pp. 423-439. Dashek, W.V., ed., CRC Press, Inc., Boca Raton.

Murphy, D.J., Cummins, I., and Ryan, A.J. (1989) Immunocytochemical and biochemical study of the biosynthesis and mobilisation of the major seed storage proteins of *Brassica napus*. Plant Physiol. Biochem. 27, 647-657.

Stalberg, K., Ellerstoem, M., Ezcurra, I., Ablov, S., and Rask, L. (1996) Disruption of an overlapping e-box-ABRE motif abolished high transcription of the napA storage-protein promoter in transgenic *Brassica napus* seeds. *Planta* 199: 515-519.

Voelker, T.A., Hayes, T.R., Cranmer, A.M., Turner, J.C., and Davies H.M. (1996) Genetic engineering of a quantitative trait: Metabolic and genetic parameters influencing the accumulation of laurate in rapeseed. *Plant J.* 9: 229-241.